EAI Endorsed Transactions
on Context-aware Systems and Applications
Research Article ICST.ORG

Face recognition based on LDA in manifold subspace

Hung Phuoc Truong[1], Tue-Minh Dinh Vo[1] and Thai Hoang Le[1,*]

[1]Faculty of Information Technology, University of Science – Vietnam National University Ho Chi Minh city, 227 Nguyen Van Cu street, HCMc, Vietnam

Abstract

Although LDA has many successes in dimensionality reduction and data separation, it also has disadvantages, especially the small sample size problem in training data because the "within-class scatter" matrix may not be accurately estimated. Moreover, this algorithm can only operate correctly with labeled data in supervised learning. In practice, data collection is very huge and labeling data requires high-cost, thus the combination of a part of labeled data and unlabeled data for this algorithm in Manifold subspace is a novelty research. This paper reports a study that propose a semi-supervised method called DSLM, which aims at overcoming all these limitations. The proposed method ensures that the discriminative information of labeled data and the intrinsic geometric structure of data are mapped to new optimal subspace. Results are obtained from the experiments and compared to several related methods showing the effectiveness of our proposed method.

Keywords: face recognition, manifold learning, semi-supervised, discriminative

Received on 10 January, 2016, accepted on 29 March, 2016, published on 02 May, 2016

doi: 10.4108/eai.2-5-2016.151209

1. Introduction

In many areas of artificial intelligence, information retrieval, and data mining, one is often confronted with intrinsically low-dimensional data lying in a very high-dimensional space. This leads one to consider methods of dimensionality reduction that allow one to represent the data in a lower dimensional space. Two of the most popular techniques for this purpose are Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

PCA is an unsupervised and an eigenvector method designed to model linear variation in high-dimensional data. PCA is guaranteed to discover the dimensionality of the subspace and produces a compact representation when the data is embedded in a linear subspace.

LDA is a supervised method. LDA searches for the project axes on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other. LDA encodes discriminating information in a linear separable space using bases are not necessarily orthogonal. When label information is available, *e.g.* for classification task, LDA can achieve significant better performance than PCA. However, recent work [4] shows that when the training dataset is small, PCA can outperform LDA. The reason is covariance matrix of each class in LDA may not be accurately estimated. There are a lot of approaches that try to improve the performance of PCA and LDA, which are [1][2][3][23][24].

Recently, a number of research efforts have shown that the face images possibly reside on a nonlinear manifold [6][10][11][16][17][18][20][21][22]. Both PCA and LDA fail to discover the underlying structure when the face images lie on a manifold since they effectively see only the Euclidean structure. There has been some interest in the problem of developing low dimensional representations through kernel based techniques for face recognition [14][15]. These methods can discover the nonlinear structure of the face images. However, they are computationally expensive, and none of them explicitly considers the structure of the manifold on which the face images possibly reside. In the meantime, some nonlinear

*Corresponding author. Email: lhthai@fit.hcmus.edu.vn

techniques have been proposed to discover the nonlinear structure of manifold, *e.g.* ISOMAP [13], LLE [6], Laplacian Eigenmap [12]. However, these nonlinear manifold learning techniques might not be suitable for face recognition since they do not generally provide a functional mapping between the high and low dimensional spaces that are valid both on and off the training data. There are a lot of approaches that try to address this issue by explicitly requiring an embedding function either linear or in reproducing kernel Hilbert space when minimizing the objective function [16][17][18]. One of the major limitations of these methods is that they fail to characterize the manifold structure of data when there are insufficient training samples. To solve this problem, many techniques have been proposed [19][20] which have significantly improved the face recognition performance. However, these recognition algorithms struggle in achieving a reliable performance under more practical environments, where facial appearances are of large variations in illumination, expression, pose. An approach based on deep neural network has been proposed [5] to learn a nonlinear embedding from a high-dimensional data space to a low-dimensional space. However, this technique is computationally expensive and hard to determine the parameters.

In reality, we usually have small part of input data labeled, along with a large number of unlabeled data. Thus, semi-supervised learning has attracted an increasing amount of attention. Two well-known algorithms are extension of Support Vector Machine [21] and graph-based learning [10][22]. Despite of their performance, it is unclear to determine the good graph.

The above described fact has shown that nonlinear manifold learning is encouraging and has good results in this field. The traditional methods such as PCA and LDA use Euclid distance in linear space, which cannot represent or simulate data in all cases. Besides that, new approaches with manifold learning is more suitable for face data because they utilize geodesic distance between pairwise points then estimates the local embeddings of data in order to update weights for each data point proportional to the distance between the point and the estimated manifold (compared to the original data). Thus, original data will be projected to another optimal subspace of manifold learning and still obtain characteristics of data. In a recent study, Thang, et al. [25] proved that data clustering in manifold learning is better than in linear subspace and proposed a scheme to overcome the limitation of the traditional Graph K-means algorithm called GKM-LC which always ensures that the number of clusters is stable in each iteration.

This paper is the extended version of our published paper [26], we propose a new semi-supervised dimensionality reduction algorithm, called Discriminative Semi-supervised Learning in Manifold subspace (DSLM). Our proposed algorithm aims to find a projection which captures not only the discriminant structure inferred from the labeled data but also the intrinsic geometrical structure inferred from the whole training data. Specifically, the training data is used to build a graph incorporating neighborhood information in which each data point is represented as a linear combination of the neighboring data points. The graph provides a discrete approximation to the local geometry of the data manifold. Figure 1 describes the general idea of proposed framework: firstly, the data is used to construct an adjacency graph; next, each edge in the graph is weighted with appropriate values, the result matrix W then used to compute a characterise matrix D; finally, by solving eigen-problem and choosing the suitable number of eigenvectors to form the projection matrix, the algorithm can find the optimal subspace for representing data.

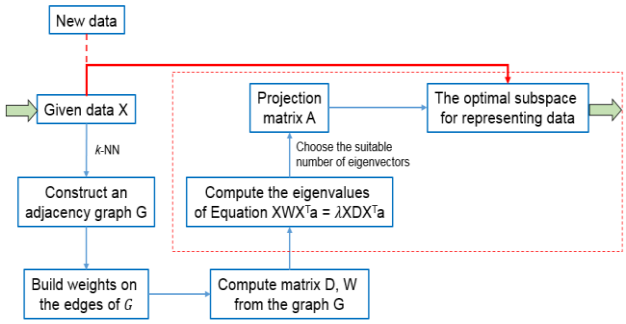

Figure 1. A general scheme of a group of extended manifold algorithms

The rest of this paper is organized as follows: The Discriminative Semi-supervised Learning in Manifold subspace (DSLM) algorithm is described in Section 2. A variety of experimental results are presented in Section 3. Section 4 discusses the effectiveness of our proposed algorithm. Finally, we provide some concluding remarks and suggestions for future work in Section 5.

2. Discriminative Semi-supervised Learning in Manifold subspace (DSLM)

2.1. The objective function

The basic idea of Linear Discriminant Analysis (LDA) is to seek directions on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other.

Suppose we have a set of n sample $X = \{\mathbf{x}_i\}_{i=1}^{n}$, $\mathbf{x}_i \in \mathbb{R}^D$ belonging to c classes (C_1, C_2, … C_n). And suppose we have the number of data points in each class is $n_{C_1}, \ldots, n_{C_c}$, $\mathbf{x}_i^{(k)}$ is the i-th point belonging to k-th class. In this space, we have some following equations for LDA algorithm:

$$\boldsymbol{\mu}^{(k)} = \frac{1}{n_{C_k}} \sum_{i=1}^{n_{C_k}} \mathbf{x}_i^{(k)} \qquad (1)$$

$$\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \tag{2}$$

$$\mathrm{S_b} = \sum_{k=1}^{c} n_{C_k}(\boldsymbol{\mu}^{(k)} - \boldsymbol{\mu})(\boldsymbol{\mu}^{(k)} - \boldsymbol{\mu})^T \tag{3}$$

$$\mathrm{S_w} = \sum_{k=1}^{c}\left(\sum_{i=1}^{n_{C_k}}\left(\mathbf{x}_i^{(k)} - \boldsymbol{\mu}^{(k)}\right)\left(\mathbf{x}_i^{(k)} - \boldsymbol{\mu}^{(k)}\right)^T\right) \tag{4}$$

The objective function of LDA is as follows:

$$\boldsymbol{a}_{opt} = \arg\max_{\boldsymbol{a}} \frac{\boldsymbol{a}^T S_b \boldsymbol{a}}{\boldsymbol{a}^T S_w \boldsymbol{a}} \tag{5}$$

where $\boldsymbol{\mu}$ is the total sample mean vector, $\boldsymbol{\mu}^{(k)}$ is the average vector of the k-th class, $\mathrm{S_b}$ is the "between-class scatter" matrix, and $\mathrm{S_w}$ is the "within-class scatter" matrix.

We define the total scatter matrix $\mathrm{S_t} = \mathrm{S_W} + \mathrm{S_b}$:

$$S_t = \sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T \tag{6}$$

Then the object function of LDA in Equation (5) is equivalent to

$$\boldsymbol{a}_{opt} = \arg\max_{\boldsymbol{a}} \frac{\boldsymbol{a}^T S_b \boldsymbol{a}}{\boldsymbol{a}^T S_t \boldsymbol{a}} \tag{7}$$

We denote the matrix $X = [X^{(1)}, \ldots, X^{(c)}]$ and the matrix W_{LDA} as

$$W_{LDA} = \begin{bmatrix} W^{(1)} & 0 & \cdots & 0 \\ 0 & W^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W^{(c)} \end{bmatrix} \tag{8}$$

where $\mathrm{W^{(k)}}$ is a $n_k \times n_k$ matrix with all elements equal to $\frac{1}{n_k}$ and $X^{(k)}$ is the data matrix of k-th class.

Without loss of generality, we assume $\boldsymbol{\mu} = 0$, which can be achieved by centering the data, *i.e.*, subtract the mean vector from all the sample vectors. Thus, we have:

$$\begin{aligned} S_b &= \sum_{k=1}^{c} n_k \boldsymbol{\mu}^{(k)}\left(\boldsymbol{\mu}^{(k)}\right)^T \\ S_b &= \sum_{k=1}^{c} n_k \left(\frac{1}{n_k}\sum_{i=1}^{n_k}\mathbf{x}_i^{(k)}\right)\left(\frac{1}{n_k}\sum_{i=1}^{n_k}\mathbf{x}_i^{(k)}\right)^T \\ S_b &= \sum_{k=1}^{c} X^{(k)} W^{(k)}\left(X^{(k)}\right)^T \\ S_b &= X W_{LDA} X^T \end{aligned} \tag{9}$$

And

$$S_t = \sum_{i=1}^{n}(\mathbf{x}_i)(\mathbf{x}_i)^T = XX^T \tag{10}$$

The object function of LDA in Equation (7) can be rewritten as [10]:

$$\boldsymbol{a}_{opt} = \arg\max_{\boldsymbol{a}} \frac{\boldsymbol{a}^T S_b \boldsymbol{a}}{\boldsymbol{a}^T S_t \boldsymbol{a}} = \arg\max_{\boldsymbol{a}} \frac{\boldsymbol{a}^T X W_{LDA} X^T \boldsymbol{a}}{\boldsymbol{a}^T X X^T \boldsymbol{a}} \tag{11}$$

When there is only one sample, LDA may be an ill-posed problem. When there is a small training set, overfitting may occur. The technique to solve those problem is regularization by introducing additional information. The optimization problem of regularized version of LDA can be written as follows [9]:

$$\max_{\boldsymbol{a}} \frac{\boldsymbol{a}^T S_b \boldsymbol{a}}{\boldsymbol{a}^T S_t \boldsymbol{a} + \alpha J(\boldsymbol{a})} \tag{12}$$

where $J(\boldsymbol{a})$ controls the learning complexity of the hypothesis family, and the coefficient α controls balance between the model complexity and the empirical loss. The regularizer term $J(\boldsymbol{a})$ provides us the flexibility to incorporate our prior knowledge on some particular applications. The key to semi-supervised learning algorithm is the prior assumption of consistency. For classification, it means nearby points are likely to have the same label [7]. For dimensionally reduction, it means nearby points will have similar low-dimensional representations. Motivated by this intuition, we take advantage of the geometric properties of manifold patches. Specifically, if the data points lying on the same patch are likely to have the same label (Figure 2).

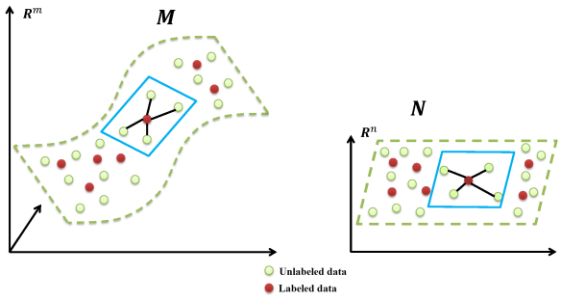

Figure 2. Data points lie on same patch This is a legend. Caption to go below figure

Suppose X is from a smooth underlying manifold of dimensionality $d \ll D$. Each data points can be reconstructed from its neighbors with appropriate weights and these weights should be the same in low-dimensional space. Let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \in \mathbb{R}^{\boldsymbol{d}}$ be the corresponded mapped data. We have the cost function of a good map [6] under appropriate constraints as:

$$\Phi(\boldsymbol{y}) = \sum_i \left(\boldsymbol{y}_i - \sum_j W_{ij}\boldsymbol{y}_j\right)^2 \tag{13}$$

which adds up the squared distances between all the data points and their reconstructions. W_i reveals the layout of the point around $\boldsymbol{x_i}$. Suppose the transformation is linear, that is, $\boldsymbol{y}_i = f(\boldsymbol{x}_i) = \boldsymbol{a}^T\boldsymbol{x}_i$. We define

$$\boldsymbol{z}_i = \boldsymbol{y}_i - \sum_j W_{ij}\boldsymbol{y}_j \tag{14}$$

which can be rewritten in vector form:

$$\boldsymbol{z} = \boldsymbol{y} - W\boldsymbol{y} = (I - W)\boldsymbol{y} \tag{15}$$

The cost function in Equation (13) can be reduced to

$$\begin{aligned}
\Phi(\boldsymbol{y}) &= \sum_i \left(\boldsymbol{y_i} - \sum_j \boldsymbol{W_{ij}}\boldsymbol{y_j} \right)^2 \\
\Phi(\boldsymbol{y}) &= \sum_i (\boldsymbol{z_i})^2 \\
\Phi(\boldsymbol{y}) &= \boldsymbol{z}^T\boldsymbol{z} \\
\Phi(\boldsymbol{y}) &= \boldsymbol{y}^T(I - W)^T(I - W)\boldsymbol{y} \\
\Phi(\boldsymbol{y}) &= \boldsymbol{a}^T X(I - W)^T(I - W)X^T\boldsymbol{a} \\
\Phi(\boldsymbol{y}) &= \boldsymbol{a}^T XMX^T\boldsymbol{a}\,.
\end{aligned} \tag{16}$$

where $M = (I - W)^T(I - W)$.

Finally, we apply the approach of LDA and use the preserving local patches cost function as a regularizer term to make the objective function of DSLM:

$$\max_{\boldsymbol{a}} \frac{\boldsymbol{a}^T S_b \boldsymbol{a}}{\boldsymbol{a}^T S_t \boldsymbol{a} + \alpha J(\boldsymbol{a})} = \max_{\boldsymbol{a}} \frac{\boldsymbol{a}^T S_b \boldsymbol{a}}{\boldsymbol{a}^T (S_t + \alpha XMX^T)\boldsymbol{a}} \tag{17}$$

Without loss of generality, we assume that the first n data points are labeled and ordered according to their labels. We use $X_l = [\mathbf{x}_1, \ldots, \mathbf{x}_l]$ to denote the labeled data matrix. We define the weight matrix $W \in \mathbb{R}^{n\times n}$ as

$$W = \begin{bmatrix} W_{LDA} & 0 \\ 0 & 0 \end{bmatrix}, \tilde{I} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \tag{18}$$

where $W_{LDA} \in \mathbb{R}^{\boldsymbol{l}\times\boldsymbol{l}}$ is defined in Equation (8), I is an identity matrix of size $l \times l$.

We have

$$S_b = X_l W_{LDA} X_l^T = XWX^T \tag{19}$$

$$S_t = X_l X_l^T = X\tilde{I}X^T \tag{20}$$

Then equation (17) describing the objective function of DSLM in can be rewritten as

$$\max_{\boldsymbol{a}} \frac{\boldsymbol{a}XWX^T\boldsymbol{a}}{\boldsymbol{a}^T X(\tilde{I} + \alpha M)X^T\boldsymbol{a}} \tag{21}$$

2.2. The algorithm

Given data set $X = \{\boldsymbol{x}_i\}_{i=1}^n$ includes labeled set $X_l = \{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^l$ belonging to c classes and ordered according to their labels, and unlabeled set $X_u = \{\boldsymbol{x_i}\}_{i=l+1}^n$. The k-th class have l_k samples, $\sum_{i=1}^c l_k = l$. The detail of the proposed method is described in Figure 3: (1) data set X is used to construct the adjacency graph G, with each image being assumed a single node of the graph; (2) we weight the edge of the graph G by using simple method (described below), the result is a weight matrix W which is used to compute matrix M; (3) we construct a special matrix $\tilde{W}$ which characterises the property of labeled data and an appropriate matrix $\tilde{I}$; (4) three feature matrices of two previous phases are combined for reconstructing optimal subspace by solving eigenvector problem.

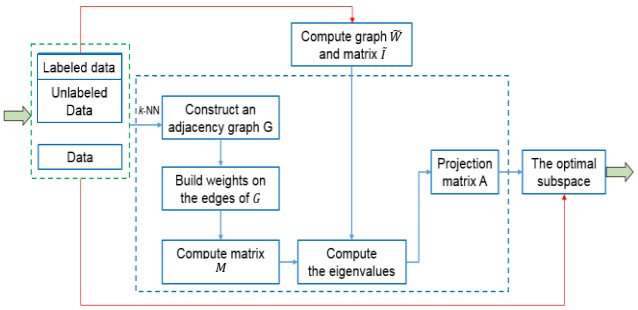

Figure 3. The proposed method (DSLM)

1. Construct the adjacency graph:

In this step, we construct the adjacency graph G of all data set X by using the k-nearest neighbors method.

2. Compute the weights:

In this step, we compute the weights on the edges of G. Let W be the weight matrix with W_{ij} having the weight of the edge from node i to node j, and 0 if there is no such edge. We define $M = (I - W)^T(I - W)$ where I is the identity matrix of size $n \times n$.

Please see [6] for details about how to compute W.

3. Construct the graph for labeled data:

In this step, we construct the weight matrix $\tilde{W} \in \mathbb{R}^{\boldsymbol{n}\times\boldsymbol{n}}$ for labeled data

$$\tilde{W} = \begin{bmatrix} W_l & 0 \\ 0 & 0 \end{bmatrix}, \tilde{I} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

where $W_l \in \mathbb{R}^{\boldsymbol{l}\times\boldsymbol{l}}$ is defined as W_{LDA} in Eqn. (8), I is an identity matrix of size $l \times l$.

4. Computing the projections:

In this step, we compute the linear projections by solving the following generalized eigenvector problem

$$X\tilde{W}X^T\boldsymbol{a} = \lambda X(\tilde{I} + \alpha M)X^T\boldsymbol{a} \tag{22}$$

It is easy to check that $\tilde{W}$ is of rank c and we will have c eigenvectors with respect to non-zero eigenvalue [8]. Let $A = [\boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_{c-1}]$ be the solution of Eqn. (22), ordered according to their eigenvalues, $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{c-1} > 0$. A is a $n \times c$ matrix. The mapping subspace is as follows

$$\mathbf{x} \to \boldsymbol{z} = A^T\mathbf{x} \tag{23}$$

5. Classification:

- Having projection matrix A which is computed using DSLM algorithm by solving Eq. (22), each testing

image T is mapped into the optimal subspace by using Eq. (23), which can be written as follows:

$$\boldsymbol{t}_i \rightarrow \boldsymbol{u}_i = A^T \boldsymbol{t}_i$$

where $\boldsymbol{t}_i$ is the feature vector of testing image T.

- Compute the Euclidean distance from the testing image to each j-th labeled image $\boldsymbol{x}'_j$ in the optimal subspace as follows:

$$d_j(T, x'_j) = \|u_i - z'_j\|_2$$
$$d_j(T, x'_j) = \|A^T t_i - A^T x'_j\|_2$$
$$d_j(T, x'_j) = \sqrt{\sum_{s=1}^{d} \left((A^T t_i)_s - (A^T x'_j)_s\right)^2}$$

- We then assign a label c to T as follows:

$$c = \arg\min_j d_j(T, x'_j) \ , j = 1,2,..,l$$

3. Experimental results

In this section, we investigate the use of our proposed approach for face recognition. We compare our DSLM algorithm with several representative dimension reduction algorithms, which include PCA, LDA, SDA [10]. PCA and LDA are the two most widely used subspace learning techniques for face recognition. SDA is the algorithm with high accuracy on semi-supervised face recognition [10].

3.1. Dataset descriptions

The YALE [27] face database contains 165 grayscale images of size 320×243 of 15 people (11 samples for person). The images demonstrate variations in lighting condition (left-light, center-light, right-light), facial expression (normal, happy, sad, sleepy, surprised and wink), and with/without glasses.

The ORL [28] face database contains 400 gray images of size 92×112 of 40 people (10 samples for person). The images were captured at different times and have different variation including expressions (open or closed eyes, smiling or non-smiling) and face details (glasses or no glasses). The images were taken with a tolerance for some tilting and rotation of face up to 20 degrees.

3.2. Data preparation and experimental settings

In all the experiments, preprocessing to locate the faces was applying. Original images were normalized (in scale and orientation) such that the two eyes were aligned at the same position. Then the facial areas were cropped into the final image for matching.

The size of each cropped image in all the experiment is 32×32 pixels, with 256 gray levels per pixel. Thus, each image can be represented by 1024-dimesional vector in image space. No further preprocessing is done. 10 images of a person in YALE and 10 images of a person in ORL are displayed in Figure 4.

Figure 4.a: ORL face database

Figure 4.b: YALE face database

Figure 4. Samples from YALE face database and ORL face database with different facial expression and details

We use the semi-supervised setting for our experiments. That is, the available training set during the training phase contains both labeled and unlabeled examples, and the testing set is not available during the training phase. In this paper, we apply nearest-neighbor classifier for its simplicity. For each person in dataset, n images are randomly selected as the training set. Among these n images, l images are randomly selected and labeled which leaves other $n - l$ images unlabeled. We average the result over 25 random split. The recognition performance is measured by the accuracy:

$$Accuracy = \frac{Nc}{Ns} \times 100\%$$

where *Nc* is the number of correctly classified test samples and *Ns* is the number of test samples

3.3. Face recognition with different dimensions

In this experiment, we fix $\alpha = 0.1$ for two methods SDA and DSLM. The number of nearest neighbors k is between 2 and 4, the recognition is carried out then. In general, the accuracy rates varies with the dimension of the face subspace. Figure 5 shows the plots of accuracy rates versus dimensionality reduction for the PCA, LDA, SDA and DSLM. The best result obtained in the optimal subspace and the corresponding dimensionality for each method are shown in Table 1. Note that the upper bound of dimensionality of SDA and DSLM is c where c is the

number of classes. When there is a single labeled training image per class, LDA cannot be applied since the within-class scatter matrix is the zero matrix. As can be seen, our DSLM algorithm performed the best for all the cases. Moreover, the optimal dimensionality obtained by DSLM, SDA and LDA is much lower than that obtained by PCA.

3.4. Face recognition with different k-nearest neighbors

The most important parameter in all of the manifold approaches which make use of the manifold structure is k-nearest neighbors. We test and compare two methods SDA and DSLM with different values of k. In this experiment, we use the ORL face database and fix $n = 7,\ l = 3, \alpha = 0.1$; k is chosen between 2 and 6; the recognition is carried out then.

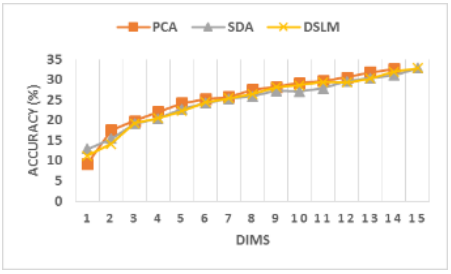

Figure 5.a: $n = 5, l = 1$

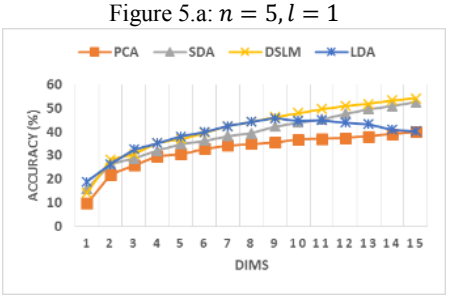

Figure 5.b: $n = 5, l = 2$

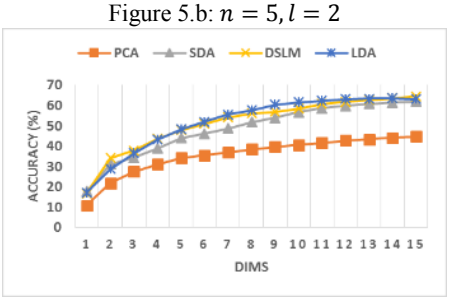

Figure 5.c: $n = 5, l = 3$

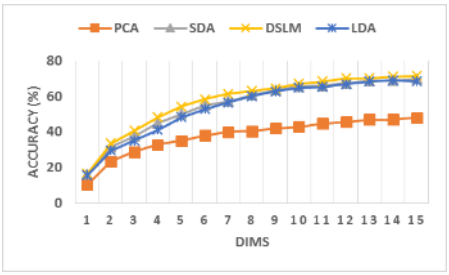

Figure 5.d: $n = 5, l = 4$

Figure 5. Accuracy rates vs. dimensionality reduction on the YALE face database

Figure 6 shows the plots of accuracy rates versus number nearest of neighbor. Table 2 shows the performance comparison of those. As can be seen, our DSLM algorithm performed better result. Moreover, the accuracy of our DSLM algorithm is stable with varying value of parameter k. It is shown that our DSLM algorithm is stability with varying size of patches on manifold.

Table 1. Performance comparisons on the YALE face database

Methods	The value of parameter l			
	$l = 1$	$l = 2$	$l = 3$	$l = 4$
PCA	32.6 (14)	43.5 (29)	50.4 (44)	54.4 (59)
LDA	-	45.8 (9)	63.6 (14)	69.1 (14)
SDA	32.8 (15)	52.5 (15)	62.1 (15)	69.7 (15)
DSLM	**32.8 (15)**	**54.2 (15)**	**64.6 (15)**	**71.6 (15)**

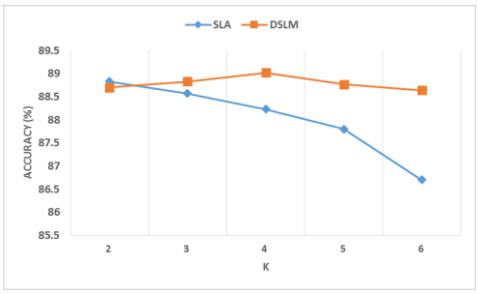

Figure 6. Accuracy rates vs. k-nearest neighbor

Table 2. Performance comparisons on the ORL face database

The parameter l	Methods	
	SDA	**DSLM**
$l = 2$	**88.83**	88.70
$l = 3$	88.57	88.83
$l = 4$	88.23	**89.02**
$l = 5$	87.80	88.77
$l = 6$	86.70	88.64

4. Discussion

It is worthwhile to high light several aspects of the proposed approach here:

1. Our proposed algorithm DSLM shares some similar properties with Semi-supervised Discriminant Analysis [10] algorithm. Both of them aim to find the optimal projection of the discriminative power of the labeled data and of the locality preserving power of manifold.

However, their graphs which discover manifold structure are totally different. Thus, their objective functions are different.

2. Some manifold learning algorithms like ISOMAP, LLE, Laplacian eigenmaps are defined only on the training data points and it is unclear how to evaluate the map for new test points. DSLM can find the optimal linear projection. Thus, this makes it fast and suitable for practical applications, e.g. face recognition.
3. DSLM can be performed and product significant results in small datasets which cannot be achieved by LDA, which can be seen as experimental results.

5. Conclusion

In this paper, we proposed a new linear dimensionality reduction algorithm called Discriminative Semi-supervised Learning in Manifold subspace. By using a graph which characterizes the locality structure of manifold data and taking advance of discriminative power of LDA method, our algorithm can make use of both labeled data and unlabeled data points to find optimal projection. Experimental results on face recognition have demonstrated the effectiveness of our algorithm.

For future works, we are interested in applying the proposed method to other graphs which characterize better the geometric properties of the dataset. On the other hand, the algorithm should be investigated in supervised mode.

Acknowledgements.

This research is the extended version of the paper with title "Discriminative semi-supervised learning in manifold subspace for face recognition" in the International Conference on Context-Aware Systems and Applications 2015 (ICCASA). This research is aslo supported by Science Research funding (T-2015.21) in University of Science, Vietnam National University - Ho Chi Minh City.

References

[1] H. Truong, T. Le. Fusion of bidirectional image matrices and 2D-LDA: an efficient approach for face recognition. In: Proc. Third Symposium on Information and Communication Technology, pp. 142-148. ACM, New York (2012).

[2] T. Le, H. Truong, H. Do, D. Vo. On approaching 2D-FPCA technique to improve image representation in frequency domain. In: Proc. Fourth Symposium on Information and Communication Technology, pp. 172-180. ACM, New York (2013).

[3] DK. Le, H. Truong, T. Le. Facial expression recognition using statistical subspace. In: IEEE International Conference on Image Processing, pp. 5981-5985. IEEE (2014).

[4] A. M. Martinez, A. C. Kak. PCA versus LDA. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no.2, pp. 228-233. (2001).

[5] Y. Huang, W. Wang, L. Wang, T. Tan. A General Nonlinear Embedding Framework Based on Deep Neural Network. In: ICPR, pp. 732-737 (2014).

[6] S. Roweis, L. K. Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science, vol. 290 (2000).

[7] D. Zhou, O. Bousquet, T. Lal, J. Weston, B. Schölkopf. Learning with local and global consistency. In: Advances in Neural Information Processing Systems 16 (2003).

[8] G. H. Golub, C. F. V. Loan. Matrix computations. Johns Hopkins University Press (1996).

[9] J. H. Friedman. Regularized discriminant analysis. In: Journal of the American Statistical Association, vol. 84, no. 405, pp. 165–175 (1989).

[10] D. Cai, X. He, J. Han. Semi-supervised Discriminant Analysis. International Conference on Pattern Recognition, pp. 1-7 (2007)

[11] Y. Chang, C. Hu, M. Turk. Manifold of Facial Expression. In: Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures. France (2003).

[12] M. Belkin, P. Niyogi. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. In: Advances in Neural Information Processing System 15. Canada (2001).

[13] J. B. Tenenbaum, Vin de Silva, Johh C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction, vol 290. Science (2000).

[14] Q. Liu, R. Huang, H. Lu, S. Ma. Face Recognition Using Kernel Based Fisher Discriminant Analysis. In: 5th FG. Washington D. C. (2002).

[15] S. Roweis, L. Saul, G. Hinton. Global Coordination of Local Linear Models. In: Advances in Neural Information Processing System 14 (2001).

[16] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, S. Lin. Graph embedding and extension: A general framework for dimensionality reduction. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 1 (2007).

[17] X. He, P. Niyogi. Locality preserving projections. In: Advances in Neural Information Processing Systems 16. MIT Press, Cambridge, MA (2003).

[18] X. Chen, J. Wei, J. Li, X. Zhang. Integrating Local and Global Manifold Structures for Unsupervised Dimensionality Reduction. In: IJCNN. China (2014).

[19] J. Lu, Y. P. Tan. Cost-sensitive subspace learning for face recognition. The IEEE Conference on Computer Vision and Pattern Recognition, pp. 2661–2666 (2010).

[20] J. Lu, Y. P. Tan, G.Wang. Discriminative multi-manifold analysis for face recognition from a single training sample per person. In: ICCV, pp. 1943–1950 (2011).

[21] M. Belkin, P. Niyogi, V. Sindhwani. Manifold regularization: A geometric framework for learning from examples. In: Journal of Machine Learning Research (2006).

[22] Goldberg AB, Li M, Zhu X. Online manifold regularization: a new learning setting and empirical study. In: Proc. 19th European conference on machine learning, pp. 393-407. Belgium (2008).

[23] H. Van, T. Le. On Discriminant Orientation Extraction Using GridLDA of Line Orientation Maps for Palmprint Identification. In: Knowledge and Systems Engineering Advances in Intelligent Systems and Computing, vol. 244, pp. 237-248 (2014).

[24] H. Van, T. Le. GridLDA of Gabor wavelet features for palmprint identification. In: Proc. Third Symposium on Information and Communication Technology, pp. 125-134 (2012).

[25] T. Ly, H. Truong, T. Le. Graph K-means with lost cluster approach for nonlinear manifold clustering. In: Proc. The

14th International Conference on Machine Learning and Cybernetics, pp. 25-30 (2015).
[26] T-M. Vo, H. Truong, T. Le. Discriminative semi-supervised learning in manifold subspace for face recognition. In: Proc. The 4th International Conference on Context-Aware Systems and Applications, (2015).
[27] P. Belhumeur, J. Hespanha, D. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 711-720 (1997).
[28] F. Samaria, A. Harter. Parameterisation of a stochastic model for human face identification. In: 2nd IEEE Workshop on Applications of Computer Vision. Florida (1994).